

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS ✓
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

1/3/1

DIALOG(R) File 351:Derwent WPI

(c) 2004 Thomson Derwent. All rts. reserv.

010516273 **Image available**

WPI Acc No: 1996-013224/ 199602

XRPX Acc No: N96-011349

Communication control method of parallel processor system - involves
performing course selection of network based on physical address

Patent Assignee: HITACHI LTD (HITA)

Inventor: SUZUKI K

Number of Countries: 002 Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 7262146	A	19951013	JP 9447245	A	19940317	199602 B
US 5808886	A	19980915	US 95403998	A	19950314	199844
JP 3402398	B2	20030506	JP 9447245	A	19940317	200330

Priority Applications (No Type Date): JP 9447245 A 19940317

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
JP 7262146	A		9	G06F-015/16	
US 5808886	A			G05B-019/18	
JP 3402398	B2		9	G06F-015/177	Previous Publ. patent JP 7262146

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-262146

(43)Date of publication of application : 13.10.1995

(51)Int.Cl.

G06F 15/16
G06F 12/10
G06F 15/173

(21)Application number : 06-047245

(71)Applicant : HITACHI LTD

(22)Date of filing : 17.03.1994

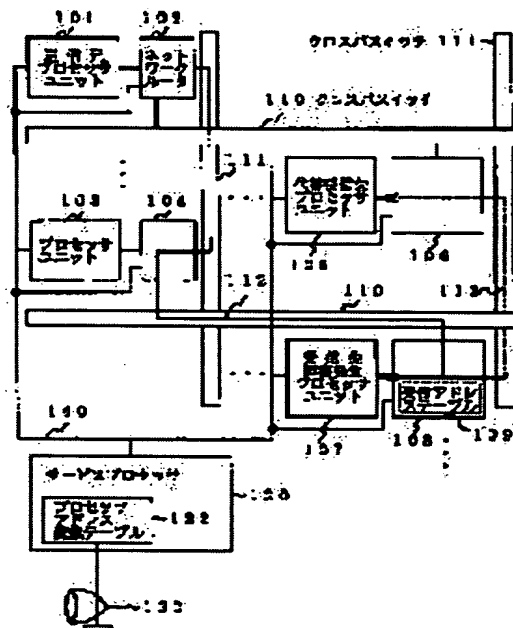
(72)Inventor : SUZUKI KAORU

(54) COMMUNICATION CONTROL METHOD FOR PARALLEL PROCESSOR SYSTEM

(57)Abstract:

PURPOSE: To quickly and efficiently perform the dynamic reconstitution processing of packet network routing.

CONSTITUTION: A transmission source processor unit 101 adds the logical address of a reception destination process unit to a packet to transmit the packet. Network routers 102 to 108 refer to a processor address conversion table 122 through a signal line 140 to obtain a physical address corresponding to the destination logical address and set a route 112 to transfer the packet to the reception destination processor unit 107. If a fault occurs in the reception destination 107, a service processor 120 changes correspondence between logical and physical addresses in the conversion table 122. As the result, a route 113 to an alternative processor unit 105 is dynamically set.



LEGAL STATUS

[Date of request for examination] 15.07.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3402398

[Date of registration] 28.02.2003

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平7-262146

(43)公開日 平成7年(1995)10月13日

(51)Int.Cl.⁶G 0 6 F 15/16
12/10
15/173

識別記号

4 7 0 H

庁内整理番号

K 7608-5B

F I

技術表示箇所

G 0 6 F 15/ 16

4 0 0 V

審査請求 未請求 請求項の数4 OL (全 9 頁)

(21)出願番号

特願平6-47245

(22)出願日

平成6年(1994)3月17日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 鈴木 薫

神奈川県秦野市堀山下1番地 株式会社日

立製作所汎用コンピュータ株式会社内

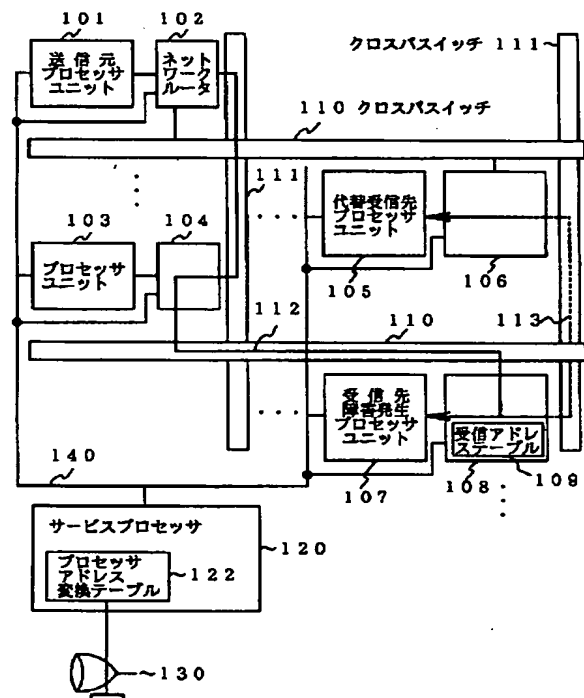
(74)代理人 弁理士 鈴木 誠

(54)【発明の名称】 並列プロセッサシステムの通信制御方法

(57)【要約】

【目的】 パケットのネットワークルーティングの動的再構成処理を高速かつ効率的に行う。

【構成】 送信元プロセッサユニット101は、受信先プロセッサユニットの論理アドレスを付加してパケットを送信する。ネットワークルータ102~108は、信号線140を介しプロセッサアドレス変換テーブル122を参照して宛先論理アドレスに対応する物理アドレスを得、ルート112を設定して、受信先プロセッサユニット107にパケットを転送する。受信先107で障害が発生すると、サービスプロセッサ120は変換テーブル122の論理/物理アドレスの対応を変更する。この結果、代替プロセッサユニット105へのルート113が動的に設定される。



1

【特許請求の範囲】

【請求項 1】 複数のプロセッサユニットが各々ネットワークルータを介してネットワーク網で結合され、プロセッサユニット間のデータ転送がパケット送受信により行われる並列プロセッサシステムにおける通信制御方法であって、

前記複数のプロセッサユニット中の任意プロセッサユニットあるいは別に設けた、システムの状態を監視する監視プロセッサに、各プロセッサユニットの論理アドレスと物理アドレスとの対応を登録したプロセッサアドレス変換テーブルを設け、

パケット送信元プロセッサユニットは自プロセッサユニットアドレス及び受信先プロセッサユニットアドレスとして論理アドレスを付加したパケットを送信し、前記パケットを受信したネットワークルータは、前記監視プロセッサに要求を出して受信先プロセッサユニットの論理アドレスに対応する物理アドレスを得、該物理アドレスにもとづいてネットワーク網の経路選択を行う、ことを特徴とする並列プロセッサシステムの通信制御方法。

【請求項 2】 複数のプロセッサユニットが各々ネットワークルータを介してネットワーク網で結合され、プロセッサユニット間のデータ転送がパケット送受信により行われる並列プロセッサシステムにおける通信制御方法であって、

各ネットワークルータに、該並列プロセッサシステムを構成する各プロセッサユニットの論理アドレスと物理アドレスとの対応を登録したプロセッサアドレス変換テーブルを設け、

パケット送信元プロセッサユニットは自プロセッサユニットアドレス及び受信先プロセッサユニットアドレスとして論理アドレスを付加したパケットを送信し、前記パケットを受信したネットワークルータは、前記プロセッサアドレス変換テーブルを参照して受信先プロセッサユニットの論理アドレスに対応する物理アドレスを得、該物理アドレスにもとづいてネットワーク網の経路選択を行う、

ことを特徴とする並列プロセッサシステムの通信制御方法。

【請求項 3】 請求項 1 もしくは 2 記載の並列プロセッサシステムの通信制御方法において、プロセッサアドレス変換テーブルの論理アドレスと物理アドレスの対応を変更して、ネットワーク網から任意プロセッサユニットを論理的に切り離すことを特徴とする並列プロセッサシステムの通信制御方法。

【請求項 4】 請求項 3 記載の並列プロセッサシステムの通信制御方法において、

各ネットワークルータに当該ネットワークルータに接続されたプロセッサユニットがパケット受信処理中であるか否かを示す識別ビットと当該パケットの送信元プロセ

2

ッサユニットの論理アドレスとの対応を登録する受信アドレステーブルを設け、

受信先プロセッサユニットの障害発生時、監視プロセッサが当該障害発生プロセッサユニットに接続されたネットワークルータの受信アドレステーブルを読み込み、前記識別 1 ビットがパケット受信処理中を示していれば、対応する論理アドレスのパケット送信元プロセッサユニットに対してパケット再送を通知することを特徴とする並列プロセッサシステムの通信制御方法。

10 【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、複数のプロセッサユニットが各々ネットワークルータを介してネットワーク網で結合され、プロセッサユニット間のデータ転送がパケット送受信によって行われる並列プロセッサシステムにおける通信制御方法に係り、特にプロセッサユニットの動的再構成を効率良く行うのに好適な並列プロセッサシステムの通信制御方法に関する。

【0002】

20 【従来の技術】 単一プロセッサの高速化の限界と、大量のデータの高速処理要求に伴って、例えば SIMD (Single Instruction Stream Multiple Data Stream; 各プロセッサが異なるデータに対して、同一の命令を実行する方式)、MIMD (Multiple Instruction Stream Multiple Data Stream; 各プロセッサが独立に個別の命令を実行する方式) など種々の方式の並列処理マシンが開発され、実用に供されている。通常、この種の並列プロセッサシステムは、数百から数万のプロセッサユニットで構成されている。

30 【0003】 従来、このような並列プロセッサシステムにおいて、プロセッサユニット間のデータ転送をパケット送受信によって行う場合、実際にプロセッサユニットが接続されている物理ネットワークアドレスを用いて、送受信プロセッサユニットを指定し、該送受信プロセッサユニット間でパケットを送受信する方法が採られている。

【0004】

【発明が解決しようとする課題】 上記した従来技術では、パケット受信動作中、または受信待ち動作中に当該プロセッサユニットで障害が発生した場合や、定期保守のために所定のプロセッサユニットをネットワーク網から切り離す必要が生じた場合、処理中のタスクを一旦中断し、プロセッサユニットの切離し、再構成が完了した後に、送信元プロセッサユニットの OS (オペレーティングシステム) は、ネットワーク構成情報を変更し、該変更された物理ネットワークアドレスを用いて、処理中の通信処理を最初からやりなおす必要があり、OS のオーバーヘッドが大きくなり、また、通信処理制御が複雑になるという問題があった。

50 【0005】 本発明の目的は、上記従来技術の問題点を

解決し、必要に応じてプロセッサユニットの動的再構成を高速かつ効率的に行うことのできる並列プロセッサシステムの通信制御方法を提供することにある。

【0006】

【課題を解決するための手段】上記目的を達成するために、本発明は、複数のプロセッサユニットが各々ネットワークルータを介してネットワーク網で結合され、プロセッサユニット間のデータ転送がパケット送受信により行われる並列プロセッサシステムにおいて、システムの状態を監視する監視プロセッサ、あるいは、複数のプロセッサユニットの各ネットワークルータに、各プロセッサユニットの論理アドレスと物理アドレスとの対応を登録したプロセッサアドレス変換テーブルを用意し、パケット送信元プロセッサユニットは宛先プロセッサユニットアドレスとして論理アドレスを付加したパケットを送信し、該パケットを受信したネットワークルータは、前記監視プロセッサに要求を出して前記論理アドレスに対応する物理アドレスを得、該物理アドレスにもとづいてネットワーク網の経路選択を行うようにしたことである。

【0007】

【作用】ネットワーク網から、例えば障害発生プロセッサユニットを切り離す必要が生じた場合、プロセッサアドレス変換テーブル内の当該プロセッサユニットの論理アドレスと対の物理アドレスを代替先プロセッサユニットの物理アドレスに変更する。これにより、障害発生プロセッサユニットのネットワーク網からの切り離しが達成されたことになる。

【0008】ネットワークルータでは、障害発生プロセッサユニットの論理アドレスが付加されたパケットを受信すると、プロセッサアドレス変換テーブルより代替先プロセッサユニットの物理アドレスを得、動的に該代替先プロセッサに対応する経路選択が実施される。また、送信元プロセッサユニットでは、パケットを再送信する場合、代替先プロセッサをまったく意識することなく、送信先プロセッサユニットアドレスとして障害発生プロセッサユニットの論理アドレスが付加されたパケットをそのまま再送信するだけでよい。

【0009】このように、本発明によれば、再構成時のソフトウェア処理が容易になり、OSのオーバーヘッドを最小限に抑えることができる。また、高速にネットワーク網からの対象プロセッサの切離し、処理中タスクの継続実行が可能になる。

【0010】

【実施例】以下、本発明の一実施例を図面を用いて具体的に説明する。図1は、本発明の一実施例のブロック構成図である。図において、101、103、105、107はプロセッサユニットであり、各々、ネットワークルータ102、104、106、108を介し、ネットワーク網を構成するXおよびYのクロスバスイッチ11

0、111に接続され、並列プロセッサシステムを構成している。109は受信パケット中の送信元プロセッサユニットの論理アドレスを一時保持する受信アドレステーブルで、便宜上、図1ではネットワークルータ108についてのみ示したが、他のネットワークルータ102、104、106にも同様に具備される。120はシステムのネットワーク網の状態を監視するサービスプロセッサ（ネットワーク監視プロセッサ）であり、保守・診断用等の専用の信号線140により、プロセッサユニット101、103、105、107及びネットワークルータ102、104、106、108と接続されている。122はサービスプロセッサ120が管理するプロセッサアドレス変換テーブルであり、各プロセッサユニット101、103、105、107の仮想的なアドレス（論理プロセッサアドレス）とネットワーク網上の物理的なアドレス（物理プロセッサアドレス）との対応を保持している。130はシステムコンソールであり、オペレータは該コンソール130を使用してプロセッサアドレス変換テーブル122の内容を設定することができる。

【0011】なお、ネットワーク網は、通常、X方向、Y方向、Z方向のクロスバスイッチで構成されているが、図1では説明を簡単にするためにX方向、Y方向の二次元のクロスバスイッチのみを示す。また、図1では4台のプロセッサユニットのみが示されているが、前述したように、この種の並列プロセッサシステムは数百から数万のプロセッサユニットで構成されている。

【0012】プロセッサユニット間のデータ転送はパケットの送受信によって行われる。いま、プロセッサユニット101を送信元、プロセッサユニット107を受信先として、まず、プロセッサユニット107が正常の場合の動作を説明する。図4に、この場合のネットワークルータの処理フローチャートを示す。

【0013】ここで、便宜上、プロセッサユニット101、ネットワークルータ102の座標位置を原点とし、右方向をX方向、下方向をY方向とする。従って、プロセッサユニット101、ネットワークルータ102の座標は(0, 0)、プロセッサユニット103、ネットワークルータ104の座標は(0, 1)、プロセッサユニット105、ネットワークルータ106の座標は(1, 0)、プロセッサユニット107、ネットワークルータ108の座標は(1, 1)となる。これらの座標値を、各プロセッサユニットの物理プロセッサアドレスとする。

【0014】送信元プロセッサユニット101は、ヘッダ部に自プロセッサユニットの論理プロセッサアドレス（送信元論理アドレス）と受信先プロセッサユニット107の論理プロセッサアドレス（宛先論理アドレス）が付加されたパケットを送信する。ネットワークルータ102は、該パケットを受信すると、ヘッダ部から宛先論

5

理アドレスを読み出し（ステップ401）、信号線140を介してサービスプロセッサ120へ、該宛先論理アドレスの付加された物理アドレス取得要求を発行することで（ステップ402）、サービスプロセッサ120から同じく信号線140を介して、該宛先論理アドレスに対応する物理アドレス（宛先物理アドレス）を得る（ステップ403）。サービスプロセッサ120でのプロセッサアドレス変換テーブル122による論理／物理アドレス変換動作については、図2により後述する。ネットワークルータ102は、自物理アドレスすなわちプロセッサユニット101の物理アドレスと取得した宛先論理アドレスとを比較し（ステップ404）、同一アドレス
10 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

【0015】パケットを受信したネットワークルータ104も、同様にしてサービスプロセッサ120から宛先物理アドレス（1，1）を得て、自分の物理アドレス（0，1）と宛先物理アドレス（1，1）とを比較し、一致しないため、その差分方向、つまりここではX方向のクロスバ112を選択してパケットを送る。

【0016】このようにして、プロセッサユニット101が送信したパケットは、112の経路でネットワークルータ108まで到達する。ネットワークルータ108でも、パケットを受信すると、同様にサービスプロセッサ120から宛先物理アドレス（1，1）を得て、自分の物理アドレス（1，1）と比較する。この結果、アドレスが一致するので、該ネットワークルータ108は、受信パケットを自プロセッサユニット107に送る（ステップ407）。この時、ネットワークルータ108は、受信パケットに付加されている送信元論理アドレスを受信アドレステーブル109に一時登録しておき、受信パケットがすべて正常にプロセッサユニット107に取り込まれると、当該送信元論理アドレスを削除する。この受信アドレステーブル109の詳細は図3により後述する。

【0017】次に、受信先プロセッサユニット107で障害が発生した場合について説明する。プロセッサユニット107は、障害が発生すると、信号線140を介してサービスプロセッサ120へ障害発生割込みを通知する。サービスプロセッサ120は、システムをフリーズ

6

（凍結）した後、プロセッサアドレス変換テーブル122を書き替え、プロセッサユニット107の論理プロセッサアドレスに対応する物理プロセッサアドレスを変更し、フリーズを解除する。ここでは、プロセッサユニット105の物理プロセッサアドレス（1，0）に変更されたとする。

【0018】いま、フリーズ時、プロセッサユニット101の送信したパケットがネットワークルータ104とネットワークルータ108の途中のクロスバ上にあったとすると、フリーズ解除で、該パケットはネットワークルータ108に到達する。ネットワークルータ108は、パケットを受信すると、上記と同様にしてサービスプロセッサ120から宛先物理アドレスを取得するが、この時、サービスプロセッサ120からは宛先物理アドレスとしてプロセッサユニット（代替受信先プロセッサユニット）105の物理アドレス（1，0）が返ってくる。ネットワークルータ108は、自分の物理アドレス（1，1）と宛先物理アドレス（1，0）とを比較し、一致しないため、その差分方向、即ち、ここではY方向のクロスバ111を選択してパケットを送信する。この結果、113の経路で、パケットは代替受信先プロセッサユニット105のネットワークルータ106に受信される。ネットワークルータ106は、サービスプロセッサ120から同様に宛先物理アドレス（1，0）を得、自分の物理アドレス（1，0）と一致するため、受信パケットを自プロセッサユニット105へ取り込む。

【0019】図2は、サービスプロセッサ120が管理するプロセッサアドレス変換テーブル122の具体的構成例、及び、該変換テーブル122を用いたプロセッサユニット間の通信および動的な再構成を説明する図である。

【0020】図1で説明したと同様にして、送信元プロセッサユニット201から送信されたパケットは、ネットワークルータ、クロスバスイッチを経由して受信先プロセッサユニット203に転送される。そのとき、各ネットワークルータは、パケットのヘッダ部に付加された宛先論理アドレスに対応する物理アドレスの取得要求をサービスプロセッサ120へ発行する。これを受けて、サービスプロセッサ120では、テーブル参照回路121を介してプロセッサアドレス変換テーブル122を参照し、宛先論理アドレス（物理プロセッサアドレス）に対応する物理アドレス（物理プロセッサアドレス）を読み出して、要求のあったネットワークルータへ返送する。

【0021】プロセッサアドレス変換テーブル122の各エントリは、プロセッサグループのクラスを指定するクラスビット、エントリの有効無効を判断する有効ビット、論理プロセッサアドレス、及び、論理プロセッサアドレスに対応する物理プロセッサアドレスで構成される。該プロセッサアドレス変換テーブル122は、サー

ビスプロセッサ120が発行するDIAG（診断）命令により初期化され、障害が発生した場合は、サービスプロセッサ120により自動的に、又は、システムコンソール130からマニュアル指示により書替えられる。

【0022】クラスビットは、サービスプロセッサ120のOSが初期化する際に、グループ分けしたプロセッサユニット群毎にシリアル番号を持っている。例えば、n台のプロセッサユニットからなる並列プロセッサシステムを、2つのグループに論理的に分けて動作させたとき、一方のプロセッサグループに例えばクラス001を割り当て、他方のプロセッサグループにクラス002を割り当てる。有効ビットは、処理参加中のプロセッサユニットに対応するエントリには“11”（使用中）が設定され、論理的に切り離されているが、動作可能なプロセッサユニットに対応するエントリには“10”（待期中）が設定され、論理的に切り離されていて、且つ動作不可であるプロセッサユニットに対応するエントリには“00”（無効）が設定されている。論理プロセッサアドレスはプロセッサユニットに任意に割り当てた仮想アドレス、物理プロセッサアドレスはネットワーク網の実際の物理アドレスを示す。

【0023】いま、送信元プロセッサユニット201の論理プロセッサアドレスを“00000001”、その物理プロセッサアドレスを“00000000”とし、受信先プロセッサユニット203の論理プロセッサアドレスを“00000009”、その物理プロセッサアドレスを“00020002”とする。また、代替受信先プロセッサユニット205の論理プロセッサアドレスを“00000004”、その物理プロセッサアドレスを“00010000”とする。

【0024】送信元プロセッサユニット201から受信先プロセッサユニット203へデータ転送する場合、プロセッサユニット201は、送信元論理アドレスとして論理プロセッサアドレス“00000001”、宛先論理アドレスとして論理プロセッサアドレス“00000009”を付加したパケットを送信する。該パケットを受信した各ネットワークルータは、宛先論理アドレス“00000009”を指定して、それに対応する物理アドレスの取得要求を、サービスプロセッサ120へ発行する。サービスプロセッサ120のテーブル参照回路121は、プロセッサアドレス変換テーブル122より論理プロセッサアドレス“00000009”に対応する論理プロセッサアドレス“00020002”を検索し、要求のあったネットワークルータへ返送する。このようにして、受信先プロセッサユニット203が正常であれば、プロセッサユニット201から送信されたパケットは、経路210でプロセッサユニット203に受信される。

【0025】一方、障害発生時におけるパケットルーティングの動的再構成は、プロセッサアドレス変換テーブル

122により以下のようにして行う。まず、プロセッサアドレス変換テーブル122について、障害プロセッサユニットと同一クラス内で有効ビットが“10”となっているエントリに対応するプロセッサユニットを、代替受信プロセッサユニットとして選択する。次に、障害発生プロセッサユニットの論理プロセッサアドレスに対応する物理プロセッサアドレスを、該選択されたプロセッサユニットの物理プロセッサアドレスに書き換え、該選択されたエントリの有効ビットを“10”から“00”に書き換えて無効にする。

【0026】例えば、受信先プロセッサユニット203に障害が発生すると、該プロセッサユニット203は、自論理プロセッサアドレス“00000009”を指定して、障害発生割込みをサービスプロセッサ120へ通知する。サービスプロセッサ120は、プロセッサアドレス変換テーブル122を参照して、同一クラス（001）内で有効ビットが“10”となっているエントリとして、論理プロセッサアドレスが“00000004”で、物理プロセッサアドレスが“00010000”であるプロセッサユニット205を代替受信プロセッサユニットとして選択する。そして、障害が発生したプロセッサユニット203のエントリの論理プロセッサアドレス“00000009”に対応する物理プロセッサアドレス“00020002”を“00010000”に書き換えるとともに、選択されたエントリの有効ビット

“10”を“00”に書き換えて無効化する。この結果、フリーズ解除後、パケットを受信したネットワークルータ204が、宛先論理アドレス“00000009”を指定して物理アドレス取得要求をサービスプロセッサ120へ発行すると、サービスプロセッサ120のテーブル参照回路121は、プロセッサアドレス変換テーブル122より論理アドレス“00000009”に対応する論理プロセッサアドレス“00010000”を検索してネットワークルータ204へ返送する。これにより、経路211を経由して、パケットが代替受信先プロセッサユニット205で受信可能となる。

【0027】図3は、各ネットワークルータが具備する受信アドレステーブルの具体的構成例、及び、その働きを説明する図である。

【0028】図3において、301、303を送信元プロセッサユニット、305を受信先プロセッサユニットとする。306は受信先プロセッサユニット305のネットワークルータであり、テーブル参照回路310、受信アドレステーブル311を具備する。受信アドレステーブル311のエントリは、当該エントリの有効・無効を示すVLDビットと、受信パケットの送信元論理アドレスが登録される論理アドレスとで構成される。図3では、ネットワークルータ306についてのみ示したが、このような受信アドレステーブルが、すべてのネットワークルータ102に内蔵されている。

【0029】いま、送信元プロセッサユニット301から送信されたパケットが受信先プロセッサユニット305のネットワークルータ306に到達すると、当該ネットワークルータ306では、テーブル参照回路310を介して、受信アドレステーブル311に、受信パケットに付加されている送信元プロセッサユニット301の論理アドレス（送信元論理アドレス）を登録し、そのエントリのVLDビットに“1”を設定する。そして、当該送信元プロセッサユニット301からのパケットがすべてプロセッサユニット305に取り込まれると（受信完了）、該ネットワークルータ306は、該当エントリのVLDビットを“0”に書き換える。テーブル参照回路310は、1エントリの登録が終了すると、エントリのポインタを次のエントリアドレスにポイントし、受信アドレステーブル306がフルになった場合、VLDビットが“0”となっている低位アドレスのエントリの先頭をポイントする。

【0030】ここで、受信先プロセッサユニット305で障害が発生した場合、該プロセッサユニット305はサービスプロセッサに障害発生割込みを通知する。これを受けてサービスプロセッサでは、図2で説明したようにプロセッサアドレス変換テーブルを書き替えてパケットルーティングの動的再構成を行う。その後、サービスプロセッサは、該障害発生プロセッサユニット305に接続されたネットワークルータ306内の受信アドレステーブル311の内容をテーブル参照回路310を通して読み取り、VLDビットが“1”すなわちパケットの受信が完了していない論理アドレスに対応する送信元プロセッサユニット（図3では301）に対してパケットの再送要求割込みを報告する。

【0031】パケットの再送要求に対して、送信元プロセッサユニット301のOSでは、受信先プロセッサユニット305の障害発生をまったく意識することなく、障害発生前と同じ条件設定のままで、パケットの送出指示を行う。この結果、該送信元プロセッサユニット301から再送されたパケットは、図2で説明したようにして、所定の代替受信先プロセッサユニットで受信される。

【0032】一方、図3において、他の送信元プロセッサユニット303からのパケットは、まだクロスバススイッチ上にあり、受信先プロセッサユニット305に到達していないので、ネットワークルータ306の受信アドレステーブル311にはその論理アドレスは登録されていない。したがって、サービスプロセッサからは該送信元プロセッサユニット303に対して、パケットの再送要求割込みを報告されない。これは、途中のパケットは、フリーズ解除後、自動的に代替受信先プロセッサユニットで受信されることによる。

【0033】また、受信アドレステーブル311内において、VLDビットが“0”、すなわち、障害発生プロ

セッサユニット305で障害発生前にパケットの受信が完了した論理アドレスに対応する送信元プロセッサユニットに対しても、サービスプロセッサからはパケットの再送要求割込みを報告されない。これは、図4で後述するように、サービスプロセッサが障害発生プロセッサユニットから代替受信先プロセッサユニットにコピーする情報に正常なパケット内容が含まれていることによる。

【0034】図5は、任意プロセッサユニットで障害が発生してから代替プロセッサユニットで処理の継続を開始するまでの、サービスプロセッサの全体的な処理フローチャートを示したものである。

【0035】ある受信先プロセッサユニットで障害が発生すると、該プロセッサユニットは障害発生割込みをサービスプロセッサに対して報告する（ステップ501）。サービスプロセッサは、障害報告割込みを受け取ると、直ちにシステム全体の動作をフリーズ（凍結）する（ステップ502）。次いで、図2で説明したように、プロセッサアドレス変換テーブルを参照して、同一クラス内で代替プロセッサユニットの対象となるエントリ（有効ビットが“10”のエントリ）を選択し（ステップ503）、該変換テーブルに登録されている障害発生受信先プロセッサユニットのエントリの物理プロセッサアドレスを、選択された代替プロセッサユニットのエントリの物理プロセッサアドレスに書き換え（ステップ504）、代替プロセッサユニットのエントリを無効化（有効ビットを“00”）する（ステップ505）。その後、障害が発生した受信先プロセッサユニットのハードウェア情報を、代替受信先プロセッサユニットにコピーする（ステップ506）。

【0036】次いで、サービスプロセッサは障害が発生した受信先プロセッサユニットに接続されているネットワークルータ内の受信アドレステーブルを読み込み（ステップ507）、VLDビットにより、該障害発生受信先プロセッサユニットがパケット受信処理中かどうか判定する（ステップ508）。そして、障害発生受信先プロセッサユニットがパケット受信処理中（VLDビットが“1”）の場合、受信アドレステーブルの当該エントリに登録されている論理アドレスの送信元プロセッサユニットに対して、パケットの再送要求割込みを報告し（ステップ509）、代替受信先プロセッサユニットに対しては、ステップ506でコピーした障害発生プロセッサユニットのハードウェア情報中の受信パケットの破棄割込みを報告して（ステップ510）、システムフリーズを解除する（ステップ511）。また、障害発生プロセッサユニットがパケット受信処理中でない場合は（VLDビットが“0”）、ステップ509、510の処理を行うことなく、システムフリーズを解除する。

【0037】図6は、本発明の第2の実施例のブロック構成図で、各ネットワークルータ602、604、606、608にそれぞれプロセッサアドレス変換テーブル

6221~6224を用意したものである。なお、図6では省略したが、該変換テーブルの他に、各ネットワークルータ602、604、606、608は図3で説明した受信アドレステーブルを持つことは第1の実施例と同様である。

【0038】サービスプロセッサ620はマスタプロセッサアドレス変換テーブル622を持っており、システム立上げ時、該マスタプロセッサアドレス変換テーブル622の内容を信号線640により各ネットワークルータ602、604、606、608に配布してプロセッサアドレス変換テーブル6221~6224を初期設定する。また、任意プロセッサユニットで障害が発生した場合、サービスプロセッサ620は、図2で説明したと同様にしてマスタプロセッサアドレス変換テーブル622を書き換え、その変更後のマスタプロセッサアドレス変換テーブル622の内容を、信号線640を介して、各ネットワークルータ602、604、606、608のプロセッサアドレス変換テーブル6221~6224に再設定する。これにより、各ネットワークルータ内のプロセッサアドレス変換テーブルの一致性が保証される。

【0039】図6の構成によれば、各ネットワークルータ602、604、606、608はパケットを受信した時、それぞれ自分のアドレス変換テーブル6221、6222、6223、6224を参照して宛先論理アドレスに対応する物理アドレスを得ることができ、一々、サービスプロセッサ620へ物理アドレス取得要求を発行する必要がなくなる。また、サービスプロセッサ620においても、該物理アドレス取得要求に一々応答する必要がなくなる。

【0040】以上、本発明の一実施例について説明したが、パケット転送路としては、クロスバスイッチに限定されるものではなく、他の任意のものを用いることができる。また、並列プロセッサシステムを構成するプロセッサユニット群の1台あるいは複数台のプロセッサユニットにサービスプロセッサの機能を兼ねさせ、サービスプロセッサが故障した場合、当該プロセッサユニットにサービスプロセッサの処理を代替させるようにしてもよい。

【0041】

【発明の効果】

(1) 本発明によれば、ネットワーク網に接続されている全てのプロセッサユニットを論理アドレスと物理アドレスで管理しているので、ネットワーク網から対象プロセッサユニットを容易に切離すことができ、動的な再構成処理を高速に行うことができ、可用性、保守性が格段

に向上する。

【0042】(2) 各プロセッサユニットの論理アドレスと物理アドレスの対応を登録したプロセッサアドレス変換テーブルは、システムの状態を監視する監視プロセッサでの一元管理、あるいは、各プロセッサユニット対応のネットワークルータでの分散管理のいずれでもよい。ここで、監視プロセッサでの一元管理は、保守が簡単であるが、ネットワークルータが一々、監視プロセッサに参加要求を発行する必要があるため、プロセッサユニットの構成数があまり多くない場合に向いている。一方、各ネットワークルータでの分散管理は、監視プロセッサへの参照要求が不要であり、プロセッサの構成数が多量の場合に有効である。

【0043】(3) 各ネットワークルータは、対応するプロセッサユニットがパケット受信処理中であるか否かを示す識別ビット、及び、当該パケットの送信元プロセッサユニットの論理アドレスの対応を管理しているので、受信中に障害があった送信元プロセッサユニットに対してのみ再送信処理を行い、パケット送信途中で、まだ受信側に到達していないパケットは、動的に再構成されても再送信処理を行うことなく、新たに割り付けられた受信先プロセッサユニットに送信される。従って、プロセッサユニットのOSの再送信処理が簡素化され、処理オーバーヘッドを大幅に削減することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施例のブロック構成図である。

【図2】プロセッサアドレス変換テーブルの構成例およびネットワークルータの動的な再構成を説明する図である。

【図3】受信アドレステーブルの構成例およびその働きを説明する図である。

【図4】ネットワークルータのパケット受信時の処理フローチャートである。

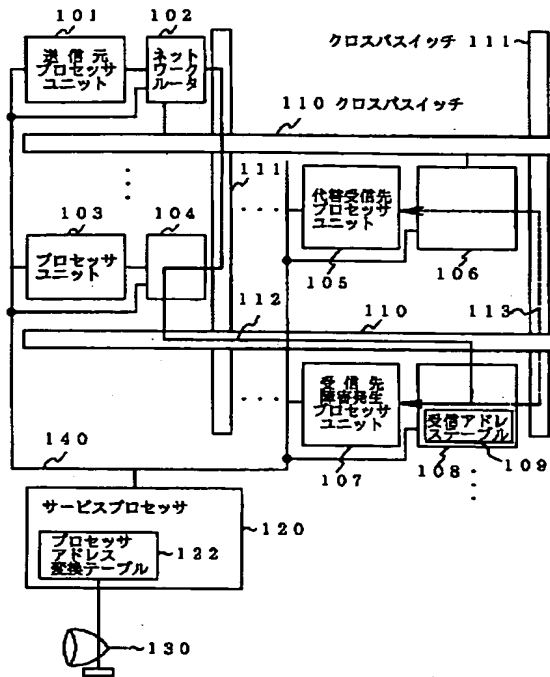
【図5】障害発生時のサービスプロセッサの動的再構成制御の処理フローチャートである。

【図6】本発明の第2の実施例のブロック構成図である。

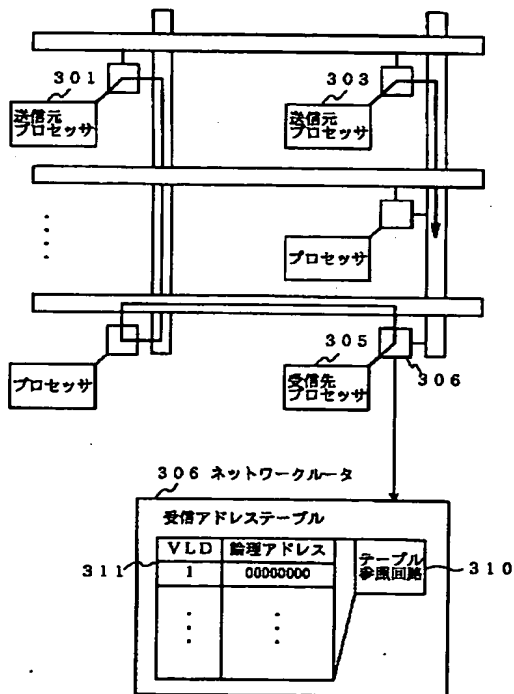
【符号の説明】

101、103、105、107 プロセッサユニット
102、104、106、108 ネットワークルータ
109 受信アドレステーブル
110、111 クロスバスイッチ
120 サービスプロセッサ
122 プロセッサアドレス変換テーブル

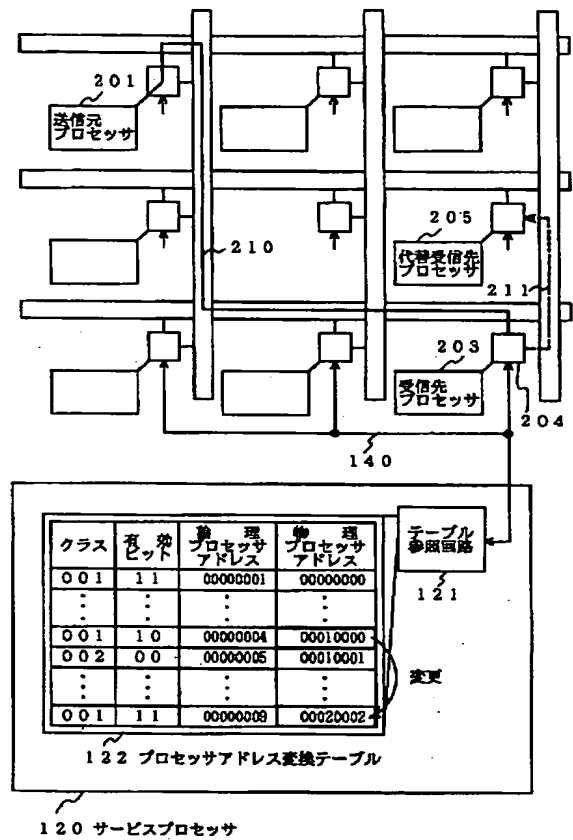
【図 1】



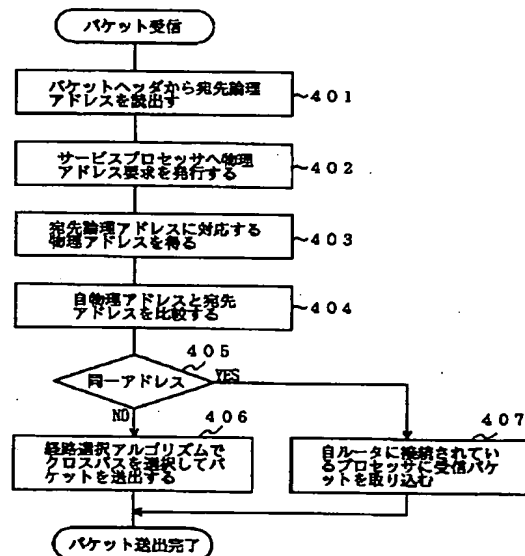
【図 3】



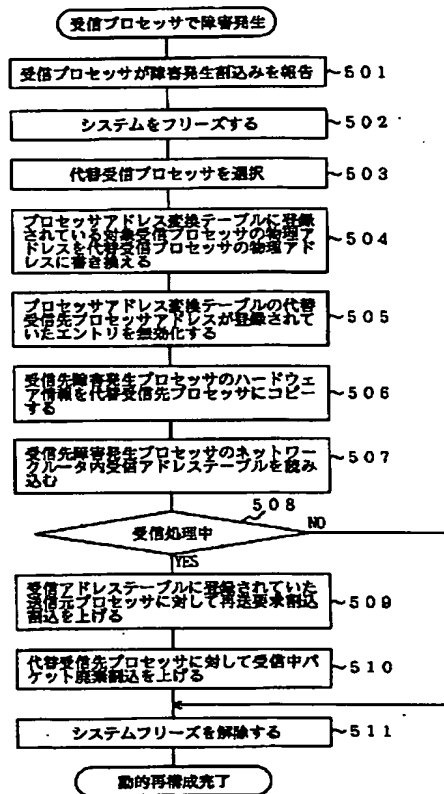
【図 2】



【図 4】



【図5】



【図6】

